dbGaP Study Release Notes

Release Notes for NHLBI TOPMed - NHGRI CCDG PMBB AF, phs001601.v2.p2
"*NHLBI TOPMed - NHGRI CCDG: Penn Medicine BioBank Early Onset Atrial Fibrillation Study*"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

May	4, 2021	Version 1 Data set release date
August	31, 2021	Version 2 Data set release date

2021-08-31
Version 2 Data set release for NHLBI TOPMed - NHGRI CCDG PMBB AF now available

This release includes updated phenotype tables and the addition of Freeze 9 whole genome sequences (WGS) and corresponding VCFs. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (IRB, PUB) (HMB-IRB-PUB)

Data Type	subjects	samples
Phenotype	2210	2253
Seq_DNA_SNP_CNV (VCFs)	2193	2236
WGS	2193	2236

For a description of SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Data Updates

1. **New Study Accession**

   NHLBI TOPMed - NHGRI CCDG PMBB AF version 1 phs001601.v1.p1 has been updated to version 2. The dbGaP accession for the current set of data is **phs001601.v2.p2**. The participant number (p#) has changed in version 2; subjects have been retired. No new subjects have been added to this study.

2. **Updated Datasets (n=3 datasets; all existing variables were updated)**

pht	version	Dataset Name
10664	2	PMBB_AF_Subject
10665	2	PMBB_AF_Sample
10666	2	PMBB_AF_Subject_Phenotypes

Molecular Data Updates

1. The molecular data is accessioned under phg001631.v1. Please see "phg001631.v1.TOPMed_CCDG_PMBB_AF_v2_frz9.sample-info.MULTI.tar.gz" folder for genotyped samples, consent status, and sample file mapping.
2. There is genotyping data in vcf 4.2 format. It is packed into a folder called "phg001631.v1.TOPMed_CCDG_PMBB_AF_v2_frz9.genotype-calls-vcf.WGS_markerset_grc38.c1.HMB-IRB-PUB.tar.gz".
3. The submitter provided QC files for this dataset are in the folder "phg001631.v1.TOPMed_CCDG_PMBB_AF_v2_frz9.genotype-qc.MULTI.tar.gz".

phg001541.v1	Freeze 8
phg001631.v1	Freeze 9

Authorized Access (Individual Level Data)

Individual level data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001601/phs001601.v2.p2

2021-05-04

Version 1 Data set release for NHLBI TOPMed - NHGRI CCDG PMBB AF now available

This release includes the addition of Freeze 8 whole genome sequences (WGS) brokered through the Sequence Read Archive (SRA), and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (IRB, PUB) (HMB-IRB-PUB)

Data Type	subjects	samples
Phenotype	2288	2253
Seq_DNA_SNP_CNV (VCFs)	421	421
WGS*	421	421

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.

For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Molecular Data

dbGaP QC steps for this release consist of checks for consistency of subject and sample IDs in phenotype and genotype components.

1. For samples and marker/enrichment-procedure info, see download components:
    a. phg001541.v1.TOPMed_CCDG_PMBB_AF.sample-info.MULTI.tar.gz
    b. phg001541.v1.TOPMed_CCDG_PMBB_AF.marker-info.MULTI.tar.gz
2. Genotypes are available in a matrix format as multi-sample vcf file(s) packed within download component(s) marked as genotype-calls-vcf. Integrity of submitted vcf files and their compatibility with PSEQ are routinely checked. Components may be divided by platform and/or population.

a. phg001541.v1.TOPMed_CCDG_PMBB_AF.genotype-calls-vcf.WGS_markerset_grc38.c1.HMB-IRB-PUB.tar.gz

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001601/phs001601.v1.p1